RGB CLOUD FREE IMAGE OF SENTINEL-2 IMAGERIES USING TEMPORAL TRANSFORMATION ALGORITHM TO IDENTIFY PADDY FIELDS IN SUBANG, WEST JAVA

Kustiyo (1), Rokhmatuloh (1), Adhi Harmoko Saputro (1), Dony Kushardono (2)

[1] University of Indonesia, Kampus UI Depok, Jakarta, 16424, Indonesia
[2] National Institute of Aeronautics and Space (LAPAN), 70 LAPAN Road, Jakarta, 13710, Indonesia
Email: kuslapan@yahoo.com; rokhmatuloh.ssi@sci.ui.ac.id; adhi@sci.ui.ac.id; dony_kushardono@lapan.go.id

KEY WORDS: Sentinel-2, paddy fields, spectral separability

ABSTRACT: Getting a single acquisition cloud free remote sensing optical satellite image in the tropical country such as Indonesia is difficult, so combining the temporal images is necessary. This research proposes the methodology for generating RGB clouds free image of optical sentinel-2 satellite images with 5 days repetition acquisition, the whole acquisition data during 2018 in Subang district of West Java was used. The topographic flattening correction and cloud masking detection were used as pre-processing steps, then the temporal transformation was applied to produce the cloud free image. The temporal transformation results of the median, quantil-15% dan quantil-85% from each band than selected to get the best RGB combination for identifying the paddy fields. Spectral separability of Paddy Fields and Non-Paddy Field using Jeffries-Matusita (JF) distance method was used to select the best RGB composites. Usually, the JF distance was used in classification process, the higher value indicated that the two objects can be separated or distinguished each other's. The results show that best natural composite base on JF distance is RGB of Minimum-Red, Minimum-Green, and Minimum-Blue with the JF distance of 0.42, 0.55, and 0.57. The best RGB True Color composite is RGB of Minimum-SWR1, Minimum-NIR1 and Minimum-Blue with the JF distance of 0.39, 0.47, and 0.49. Reviewing the results using visual interpretation shows that combining the natural color and true color composite can be used to identify paddy and non-paddy fields precisely. Generally, the paddy field is identified as blue and green color combination from the true color and natural color. Fishpond and paddy fields can be well separated, fishpond identified as blue color in true color composite and identified as white color in natural color composite. In north part of Subang the natural color composite separates the fishpond and paddy fields better than true color composite, but the sugar plantation is separate better in true color composite. Finally, this cloud free image results can be used to update the present paddy field map. This kind of Sentinel-2 data can be provided annually, so the paddy field can be monitor periodically.

1. INTRODUCTION

In the tropical region such as Indonesia, getting a such cloud free remote sensing image is a challenging. At least a tree years composite of Landsat-8 was used to create cloud free images of Indonesia, the more frequent acquisition or higher repetition acquisition was needed to get annual cloud free image of Indonesia (Kustiyo, Purwanto, Wijaya, & Roswintiarti, 2018). The Sentinel-2 satellite with two satellite constellations monitor earth object for every 5 days (SUHET, 2015). It's more than tree time frequent compare with Landsat-8. In every 15 days, there are 3 images of Sentinel-2 in the certain locations, but there is one image for Landsat-8 image. The Sentinel-2 image has a better pixel resolution and temporal period compare with Landsat-8, it could be used to identify paddy field area better.

Right now, Indonesia have one map policy (Pemerintah_Indonesia, 2016), any government institution in Indonesia has to use the one map, especially paddy field map. This map is a consensus map, it means that all related institution already agrees with one map. The map needs to be updated periodically, so it needs a cloud free image. This research proposes the methodology for generating RGB clouds free image of optical sentinel-2 satellite images with 5 days repetition acquisition, the whole acquisition data during 2018 in Subang district of West Java was used. The Jeffries Matusita distance separability index is used to select the best RGB composite for detecting the paddy field.

2. DATA AND METHODS

2.1. Data and Research Site

Subang district, north part of West Java was used as research site (Figure 1a). Mainly, the land-use of Subang is dominated with paddy field (Figure 1b). In the north part of Subang the area is flat, it is dominated by irrigated paddy fields, the water source was supported by Jatiluhur reservoir. In the south part is mountain area, where the water is available during whole year (BPS Subang, 2019). This area usually was planted by paddy two or three times a year.

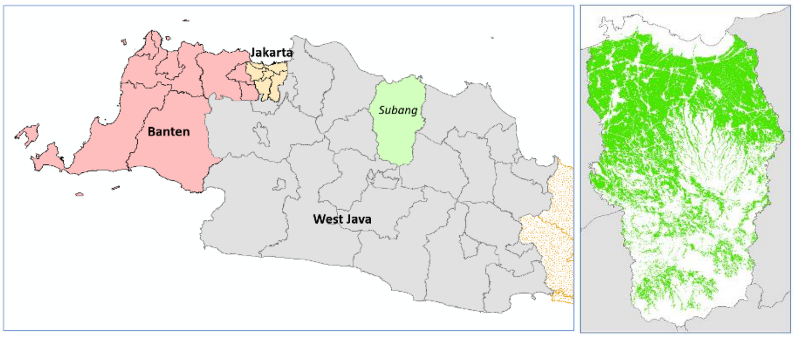


Figure 1: (a) Subang district (b) Paddy Field Map of Subang

Sentinel -2 is a multispectral satellite developed by the European Space Agency (ESA), it acquires 13 spectral bands with the spatial resolution of 10m, 20m and 60m depending on the band, as illustrated in the Table.1. (SUHET, 2015). This research use band 2, band 3, band 4, band 8, band 8A, band 11 and band 12. It was used to produce RGB natural color (Red, Green, Blue) and RGB true color (SWIR, NIR, Red). All Sentinel-2 Top-of-Atmosphere (ToA) Reflectance collection in 2018 from Google Earth Engine were used. The data is available in Level-1C orthorectified level.

Table 1: Band spectral of Sentinel-2

Sentinel-2 Bands	Central Wavelength [micrometers]	Resolution [meters]
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.49	10
Band 3 - Green	0.56	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.74	20

Band 7 - Vegetation Red Edge	0.783	20
Band 8 – NIR (NIR1)	0.842	10
Band 8A - Vegetation Red Edge (NIR2)	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 – SWIR (SWR1)	1.61	20
Band 12 – SWIR (SWR2)	2.19	20

2.2. Methodology

The processing flow was shown in Figure 2. The Pre-processing and Temporal transformation was processed on-line using GEE. The results were downloaded and then processed off-line.

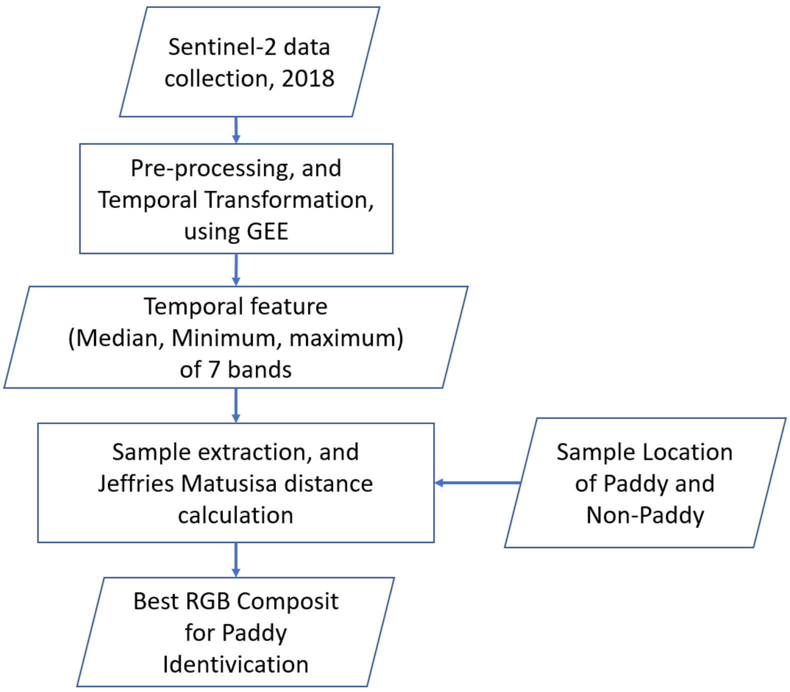


Figure 2. Processing Flow

Pre-processing:
The input data from ToA (Top of Atmospheric) Sentinel-2 collection from GEE is already in scaled reflectance. Additional processing was applied to correct the topographic effect. The are many algorithms for topographic correction, such as c-correction, minieart, cos-teta, and Illumination Correction (Hudjimartsu et al., 2018). This topographic correction uses the Illumination Correction algorithm.
Temporal transformation was applied to produce the cloud free image. The temporal transformation extracts the temporal statistical for each band (Kustiyo et al., 2018), the reflectance statistic properties of median, quantil-15% (minimum), and quantil-85% (maximum) was chosen. The final pixel resolution was saved to 20x20 meter.

Data Sample:
In general, sampling methods can be divided into two types, (1) Probability or random sampling,

and (2) Non- probability or non- random sampling. Probability sampling means that every item in the population has an equal chance to be included in sample. One of Probability sampling method is systematic sampling. The advantage of this sampling technique is its simplicity (Taherdoost, 2018). This research used systematic grid sampling to select location sampling, the distance between each sample around 500 meters. Each location was attributed by paddy field and Non-Paddy field classes. The combinations of one map paddy fields and high-resolution imagery of SPOT-6/7 in 2018 was used to decide the attribute of each sample.

Selecting the best RGB for identify paddy field:
Spectral separability of paddy fields and non-paddy field using Jeffries-Matusita (JM) distance method was used to select the best RGB composites. Usually, the JF distance was used in classification process (Wang, Qi, & Liu, 2018) (Dabboor, Howell, Shokr, & Yackel, 2014) (Sen, Goswami, & Chakraborty, 2019), the higher value indicated that the two objects can be separated or distinguished each other's.

The JM distance of paddy field and non-paddy field was calculated on seven bands and three temporal statistic parameters, so there are 7x3 or 21 JF distance. The JF distance calculated based on the formula follows:

$$J_{ij} = (1 - e^{-B_{ij}}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$B_{ij} = \frac{1}{8}\left(\mu_I - \mu_J\right)^2 \frac{2}{v_I^2 + v_J^2} + \frac{1}{2}\,\mathrm{LN}\left(\frac{v_I^2 + v_J^2}{2 v_I v_J}\right) \ldots\ldots\ldots\ldots (2)$$

where *Jij* is the Jeffries Matusita distance, *Bij* is the Bhattacharyya distance, μi and μj are the means, and *vi* and *vj* are the variance of adjacent segments *i* and *j*, respectively. The JM distance is a widely used measure of the spectral separability distance between the two class density functions.

3. RESULTS AND DISCUSSION

3.1 Pre-Processing

The final pre-processing steps of this research is the temporal statistic transformation results. The results of the three-statistic parameter from seven bands are in Figure 3. All picture is cloud free image; it was derived from cloud masking images. All the maximum image is brighter then minimum and median image, and minimum image are darker. The higher difference contras between minimum, median and maximum indicate the dynamic change during 2018. The same contras of minimum, median and maximum indicates the stable reflectance, it means that the object is not changed.

	Blue	Green	Red	NIR 1	NIR 2	SWR 1	SWR 2
Median							

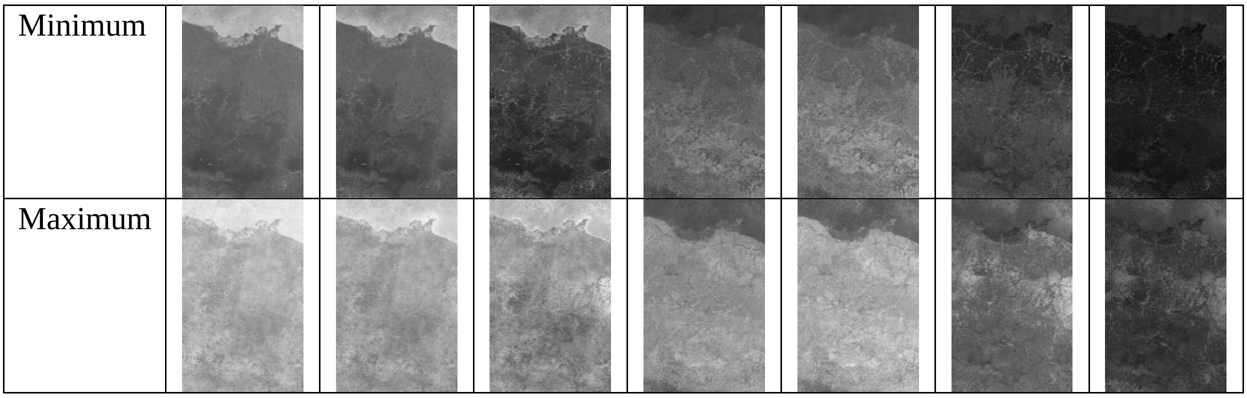


Figure 3: Temporal statistic results from Sentinel-2 data in 2018

Using the all image in Figure 3, there are many possibility RGB combinations of natural color (blue, green, red), and true color (SWR, NIR, Visible band). The combination can be come from median only, minimum only, and maximum only, or combine of median, minimum, and maximum. The Figure 3 shows that image statistic temporal results of median, minimum and maximum from band Blue and Green is similar, also between band NIR-1 and NIR-2. Visually, the contras of all visible band (blue, green, and red) are similar for maximum layer. The SWR-1 is brighter than SWR-2.

3.2 Sample selection

The Figure 4a shows the sample locations and their attribute, the total sample is 10939, 4138 (38%) are paddy field, and 6801(62%) locations is non-paddy fields. The reflectance statistic (mean and standard deviation) of each visible, NIR, and SWR band were calculated (Figure 4b), then the mean distance of paddy and non-paddy was estimated to show the separability of that two classes. Figure 4c shows that for visible bands the median image has the highest mean distance, but for NIR and SWR-1 the highest mean distance is minimum image. The SWR-2 the median and minimum image have the similar mean distance.

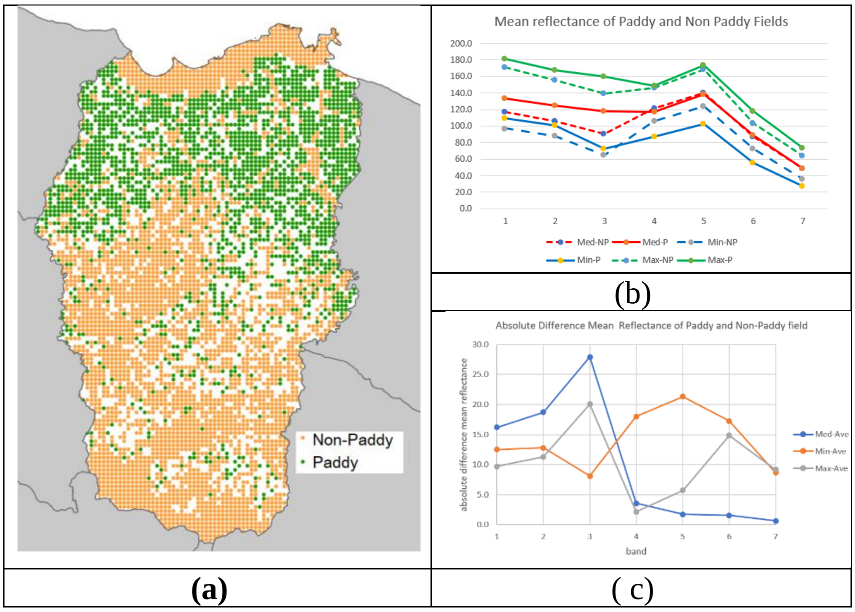


Figure 4: (a) Training sample of paddy and non-paddy field in Subang district; (b) mean reflectance each class; and (c) mean distance of two classes

3.3 Jeffries Matusita Distance

JF distance was used to evaluate two classes separability, it was derived from combination mean and variance from each class. The result was shown in Figure 5, the best or the maximum JF distance for all bands (visible, NIR, and SWIR) are temporal transformation result of minimum, and the worst JF distance are temporal transformation result of maximum for band visible and NIR, but for band SWIR the worst JF distance are temporal transformation result of median. The JF distance for band NIR (band 4 and band 5) are similar for minimum and median temporal transformation results.

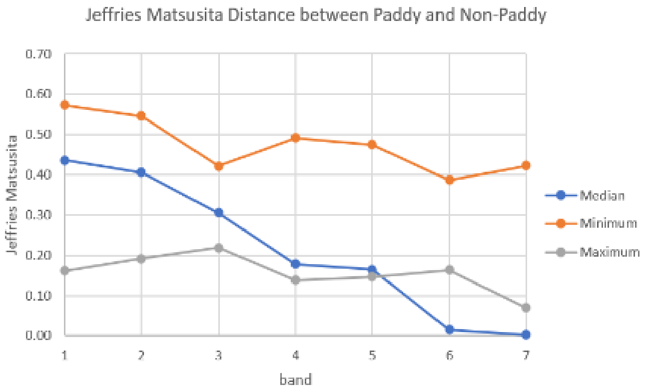


Figure 5: Jeffry Matusita distance between Paddy field and Non-Paddy fields

3.4 RGB Cloud free mosaic

The are many possible combination results for true colour and natural colour combinations from 3 annual transformation of minimum, median and maximum. The natural colour combinations are combination of Red, Green, Blue, and the true-colour bands are SWR, NIR, Visible. There are two SWR, two NIR and 3 visible bands. The SWIR-1, NIR (band 4), and visible (band blue) were selected for true colour combination, so there are 9 composites, just the 3 composite selected, it can be seen in Figure 6.a 6.b and 6.c for natural colour composite, and in Figure 6.d 6.e and 6.f for true colour composite.

The best natural colour composites form Figure 6 is Figure 6.a. Paddy field area can be separated with other land cover easily. Most of paddy area is in the same colour, the paddy area is lighter than others. In the Figure 6.b, the paddy area is in difference colour, some in red dominated by land, and some in blue dominated by water. The Figure 6.c. is the worst natural composite, because of cloud.

The true colour composite from all minimum temporal transformation in Figure 6.c shows that most of all paddy field is in blue, most of the paddy field is in one colour. The others colour composites in Figure 6.b. and 6.c. show that the paddy field are in the difference colour. It means that the Figure 6.b. and 6.c. are better used to separate the type of paddy fields but not for separate paddy filed and non-paddy fields.

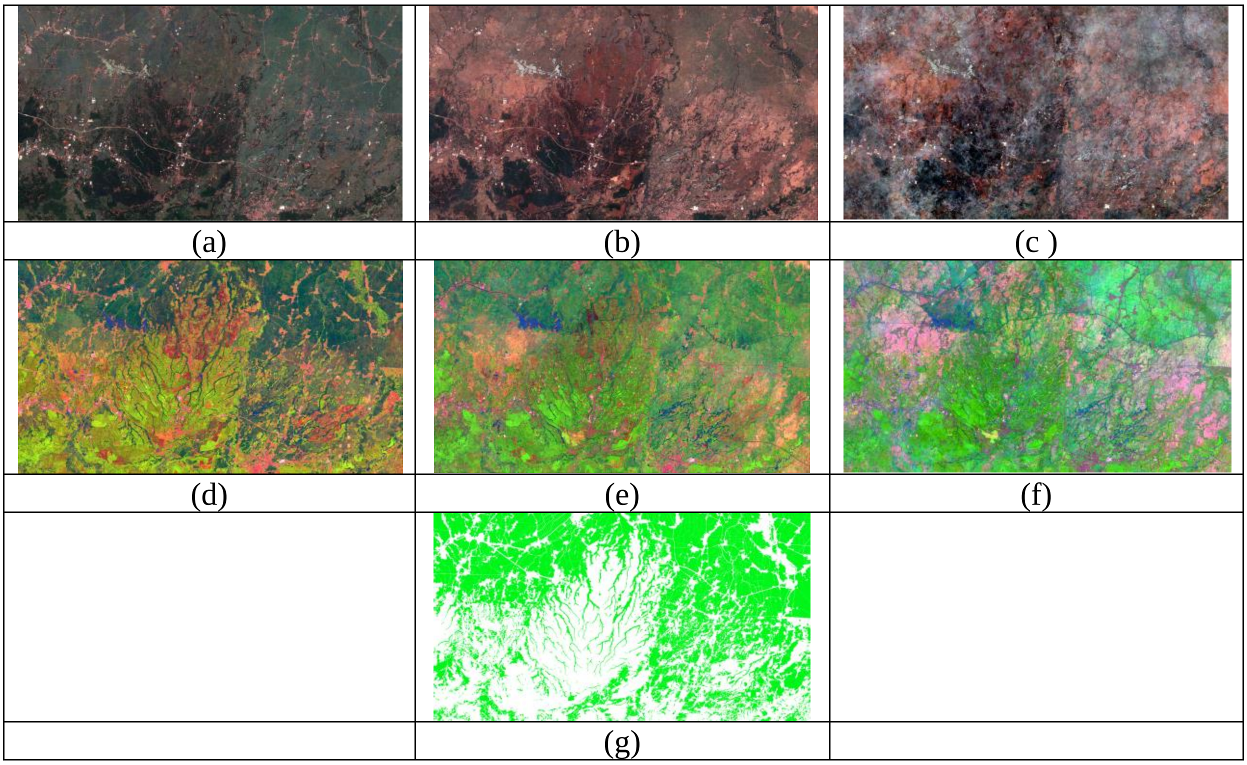


Figure 6: RGB annual transformation results cloud free mosaic of natural colour Red Green Blue and true colour SWR-1,NIR-1, Blue; (a): minimum-Red, minimum-Green, minimum-Blue; (b): median-Red, median-Green, median-Blue; (c) maximum-Red, maximum-Green, maximum-Blue; (d): minimum-SWR1, minimum-NIR1, minimum Blue; (e): median-SWR1, median-NIR1, median-Blue; (f) maximum-SWR1, maximum-NIR1, maximum-Blue; and (g) paddy fields map

Comparison of natural and true color composite in identifying paddy fields was shown in Figure 7. The four location were chosen in any difference topographic condition, location 1 is the flat area in the north part of Subang close to the beach, location 2 also the flat area, but in the center of Subang. Location 3 and location 4 are in the terrain area, topographic condition in the location 3 is less than location 4.

In the location 1, the paddy field is identified easier by natural color, the fish pond in the north part is well separated with the paddy field in the natural color, but in the true color the fish pond and paddy field can not be separated easily. This cause by the characteristic of visible band that responsive to the water cover. The fish pond is covered by water annually, but the paddy field covered by water in the time of plantation.

In the location 2, in the center of picture is fish pond, it was cover by water in every month, this fish pond can be separated well with paddy field in the both of true color and natural color composite. The fish pond is in blue color in the true color composite, but in the natural color the fish pond is white. Overlay with paddy field map from national one map, it can be seen that some area was change from paddy field to fish pond.

In the location 3, the paddy field is separated well with sugar-plantation in true color composite. The paddy area in the in lowest area, where the water is enough for paddy growth. The sugar-plantation is in the higher topographic. The border of paddy field and sugar-plantation is identified clearly in true color composite. From this location, it can be explained that the NIR and SWIR band can be use to separate paddy field and sugar-plantation.

In the location 4, there is many types of paddy fields, each location gets the water supply differently, it depends on topographic condition. Mainly, there are two difference color in true color for paddy field, but for natural color there is one color for paddy area.

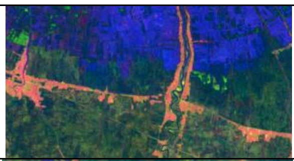
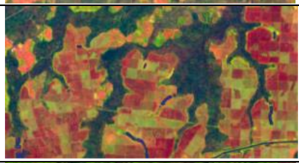
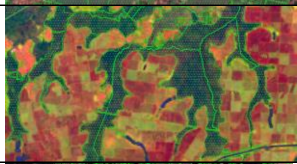
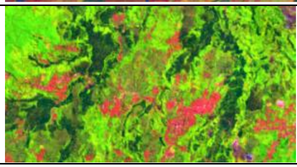
Lacation	RGB true colour	RGB natural colour	Paddy fields
Loc-1			
Loc-2			
Loc-3			
Loc-4			
	(a)	(b)	(c)

Figure 7: Samples of paddy field area from flat area (location 1), to mountainig area (location 4) for natural color (a) and true color composite (b)

3.5 Discussion

This research used the JF distance to select the best RGB composite in detecting paddy fields. Usually the JF distance is used to select the band to process classification. For classification, the selected band can be any band with the maximum JF distance. In this research the selected band was chosen to generate RGB natural color and RGB true color. The RGB natural color, the selected band must be Red, Green and Blue, but the RGB true color the selected band must be SWIR-1, NIR (NIR-1 or NIR-2) and Visible (Red, Green or Blue). The alternative selection in this RGB composite is the temporal transformation results, there is three alternatives minimum (quamtite-15%), median or maximum (quantile-85%). The minimum value is quantile 15 %, it will delete the real minimum reflectance that come from dark object such as cloud shadow. The maximum reflectance is quantil-85%, it will eliminate the real maximum reflectance such as cloud.

Base on JF distance, the best natural color combination is RGB of minimum-Red, minimum-Green, and minimum-Blue with the JF distance of 0.42, 0.55, and 0.57. The best true color combination is RGB of minimum- SWIR1, minimum- NIR1, and minimum- Blue with the JF distance of 0.39, 0.47, and 0.49. To identify paddy field, the both two RGB composite have to be combined, each combination have their strengthen in differentiate between paddy field and other land cover. In differentiating the paddy field and fish pond, the RGB natural composite is better, but for differentiate paddy field and sugar-plantation the RGB true color is better. In the terrain area the combination from both composites is better.

Finally, this cloud free image results can be used to update the present paddy field map. This kind of Sentinel-2 data can be provided annually, so the paddy field can be monitor periodically.

4. CONCLUSIONS

The best Natural Composite base on JF distance is RGB of Minimum Red, Minimum Green, and Minimum Blue, and the best RGB True Color composite is RGB of Minimum SWR1, Minimum NIR and Minimum Blue. Reviewing the results using visual interpretation shows that combining the natural color and true color composite can be used to identify paddy and non-paddy fields precisely.

References

BPS Subang. (2019). *Subang Dalam Angka*. BPS Kabupaten Subang.

Dabboor, M., Howell, S., Shokr, M., & Yackel, J. (2014). The Jeffries–Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data. *International Journal of Remote Sensing*, *35*(19), 6859–6873. https://doi.org/10.1080/01431161.2014.960614

Hudjimartsu, S., Prasetyo, L., Setiawan, Y., Ikbal, W., Hudjimartsu, S., Prasetyo, L., … Suyamto, D. (2018). preprocessing for national forest monitoring system, (August). https://doi.org/10.1117/12.2326100

Kustiyo, Purwanto, J., Wijaya, A., & Roswintiarti, O. (2018). A Statistical Approach for Cloud Free Mosaic of Landsat-8 Imageries. *Asian Conference on Remote Sensing ACRS*, 2111–2116.

Pemerintah_Indonesia. (2016). *Peraturan Presiden Republik Indonesia Nomor 9 Tahun 2016 Tentang Percepatan Pelaksanaan Kebijakan Satu Peta Pada Tingkat Ketelitian Peta Skala 1:50.000*.

Sen, R., Goswami, S., & Chakraborty, B. (2019). Jeffries-Matusita distance as a tool for feature selection. In *2019 International Conference on Data Science and Engineering (ICDSE)* (pp. 15–20). https://doi.org/10.1109/ICDSE47409.2019.8971800

SUHET. (2015). *Sentinel-2 User Handbook*. European Space Agency.

Taherdoost, H. (2018). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electronic Journal*, (September). https://doi.org/10.2139/ssrn.3205035

Wang, Y., Qi, Q., & Liu, Y. (2018). Unsupervised segmentation evaluation using area-weighted variance and Jeffries-Matusita distance for remote sensing images. *Remote Sensing*, *10*(8). https://doi.org/10.3390/rs10081193